

McMillan, James H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8). Retrieved September 24, 2011 from <http://PAREonline.net/getvn.asp?v=7&n=8> . This paper has been viewed 180,808 times since 9/23/2000.

## **Fundamental Assessment Principles for Teachers and School Administrators**

[James H. McMillan](#)

Virginia Commonwealth University

While several authors have argued that there are a number of "essential" assessment concepts, principles, techniques, and procedures that teachers and administrators need to know about (e.g. Calfee & Masuda, 1997; Cizek, 1997; Ebel, 1962; Farr & Griffin, 1973; Fleming & Chambers, 1983; Gullickson, 1985, 1986; Mayo, 1967; McMillan, 2001; Sanders & Vogel, 1993; Schafer, 1991; Stiggins & Conklin, 1992), there continues to be relatively little emphasis on assessment in the preparation of, or professional development of, teachers and administrators (Stiggins, 2000). In addition to the admonitions of many authors, there are established professional standards for assessment skills of teachers (*Standards for Teacher Competence in Educational Assessment of Students* (1990), a framework of assessment tasks for administrators (Impara & Plake, 1996), the Code of Professional Responsibilities in Educational Measurement (1995), the Code of Fair Testing Practices (1988), and the new edition of *Standards for Educational and Psychological Testing* (1999). If that isn't enough information, a project directed by Arlen Gullickson at The Evaluation Center of Western Michigan University will publish standards for evaluations of students in the near future.

The purpose of this article is to use suggestions and guidelines from these sources, in light of current assessment demands and contemporary theories of learning and motivation, to present eleven "basic principles" to guide the assessment training and professional development of teachers and administrators. That is, what is it about assessment, whether large-scale or classroom, that is fundamental for effective understanding and application? What are the "big ideas" that, when well understood and applied, will effectively guide good assessment practices, regardless of the grade level, subject matter, developer, or user of the results? As Jerome Bruner stated it many years ago in his classic, *The Process of Education*: "...the curriculum of a subject should be determined by the most fundamental understanding that can be achieved of the underlying principles that give structure to that subject." (Bruner, 1960, p.31). What principles, in other words, provide the most essential, fundamental "structure" of assessment knowledge and skills that result in effective educational practices and improved student learning?

### **Assessment is inherently a process of professional judgment.**

The first principle is that professional judgment is the foundation for assessment and, as such, is needed to properly understand and use all aspects of assessment. The measurement of student performance may seem "objective" with such practices as machine scoring and multiple-choice test items, but even these approaches are based on professional assumptions and values. Whether that judgment occurs in constructing test questions, scoring essays, creating rubrics, grading participation, combining scores, or interpreting standardized test scores, the essence of the process is making professional interpretations and decisions. Understanding this principle helps teachers and administrators realize the importance of their own judgments and those of others in evaluating the quality of assessment and the meaning of the results.

### **Assessment is based on separate but related principles of measurement evidence and evaluation.**

It is important to understand the difference between measurement evidence (differentiating degrees of a trait by description or by assigning scores) and evaluation (interpretation of the description or scores). Essential measurement evidence skills include the ability to understand and interpret the meaning of descriptive statistical procedures, including variability, correlation, percentiles, standard scores, growth-scale scores, norming, and principles of combining scores for grading. A conceptual understanding of these techniques is needed (not necessarily knowing how to compute statistics) for such tasks as interpreting student strengths and weaknesses, reliability and validity evidence, grade determination, and making admissions decisions. Schafer (1991) has indicated that these concepts and techniques comprise part of an essential language for educators. They also provide a common basis for communication about "results," interpretation of evidence, and appropriate use of data. This is increasingly important given the pervasiveness of standards-based, high-stakes, large-scale assessments. Evaluation concerns merit and worth of the data as applied to a specific use or context. It involves what Shepard (2000) has described as the systematic analysis of evidence. Like students, teachers and administrators need analysis skills to effectively interpret evidence and make value judgments about the meaning of the results.

### **Assessment decision-making is influenced by a series of tensions.**

Competing purposes, uses, and pressures result in tension for teachers and administrators as they make assessment-related decisions. For example, good teaching is characterized by assessments that motivate and engage students in ways that are consistent with their philosophies of teaching and learning and with theories of development, learning and motivation. Most teachers want to use constructed-response assessments because they believe this kind of testing is best to ascertain student understanding. On the other hand, factors external to the classroom, such as mandated large-scale testing, promote different assessment strategies, such as using selected-response tests and providing practice in objective test-taking (McMillan & Nash, 2000). Further examples of tensions include the following.

- Learning vs auditing
- Formative (informal and ongoing) vs summative (formal and at the end)
- Criterion-referenced vs norm-referenced
- Value-added vs. absolute standards
- Traditional vs alternative
- Authentic vs contrived
- Speeded tests vs power tests
- Standardized tests vs classroom tests

These tensions suggest that decisions about assessment are best made with a full understanding of how different factors influence the nature of the assessment. Once all the alternatives understood, priorities need to be made; trade-offs are inevitable. With an appreciation of the tensions teachers and administrators will hopefully make better informed, better justified assessment decisions.

### **Assessment influences student motivation and learning.**

Grant Wiggins (1998) has used the term 'educative assessment' to describe techniques and issues that educators should consider when they design and use assessments. His message is

that the nature of assessment influences what is learned and the degree of meaningful engagement by students in the learning process. While Wiggins contends that assessments should be authentic, with feedback and opportunities for revision to improve rather than simply audit learning, the more general principle is understanding how different assessments affect students. Will students be more engaged if assessment tasks are problem-based? How do students study when they know the test consists of multiple-choice items? What is the nature of feedback, and when is it given to students? How does assessment affect student effort? Answers to such questions help teachers and administrators understand that assessment has powerful effects on motivation and learning. For example, recent research summarized by Black & Wiliam (1998) shows that student self-assessment skills, learned and applied as part of formative assessment, enhances student achievement.

### **Assessment contains error.**

Teachers and administrators need to not only know that there is error in all classroom and standardized assessments, but also more specifically how reliability is determined and how much error is likely. With so much emphasis today on high-stakes testing for promotion, graduation, teacher and administrator accountability, and school accreditation, it is critical that all educators understand concepts like standard error of measurement, reliability coefficients, confidence intervals, and standard setting. Two reliability principles deserve special attention. The first is that reliability refers to scores, not instruments. Second, teachers and administrators need to understand that, typically, error is underestimated. A recent paper by Rogosa (1999), effectively illustrates the concept of underestimation of error by showing in terms of percentile rank probable true score hit-rate and test-retest results.

### **Good assessment enhances instruction.**

Just as assessment impacts student learning and motivation, it also influences the nature of instruction in the classroom. There has been considerable recent literature that has promoted assessment as something that is integrated with instruction, and not an activity that merely audits learning (Shepard, 2000). When assessment is integrated with instruction it informs teachers about what activities and assignments will be most useful, what level of teaching is most appropriate, and how summative assessments provide diagnostic information. For instance, during instruction activities informal, formative assessment helps teachers know when to move on, when to ask more questions, when to give more examples, and what responses to student questions are most appropriate. Standardized test scores, when used appropriately, help teachers understand student strengths and weaknesses to target further instruction.

### **Good assessment is valid.**

Validity is a concept that needs to be fully understood. Like reliability, there are technical terms and issues associated with validity that are essential in helping teachers and administrators make reasonable and appropriate inferences from assessment results (e.g., types of validity evidence, validity generalization, construct underrepresentation, construct-irrelevant variance, and discriminant and convergent evidence). Of critical importance is the concept of evidence based on consequences, a new major validity category in the recently revised *Standards*. Both intended and unintended consequences of assessment need to be examined with appropriate evidence that supports particular arguments or points of view. Of equal importance is getting teachers and administrators to understand their role in gathering and interpreting validity evidence.

**Good assessment is fair and ethical.**

Arguably, the most important change in the recently published *Standards* is an entire new major section entitled "Fairness in Testing." The *Standards* presents four views of fairness: as absence of bias (e.g., offensiveness and unfair penalization), as equitable treatment, as equality in outcomes, and as opportunity to learn. It includes entire chapters on the rights and responsibilities of test takers, testing individuals of diverse linguistic backgrounds, and testing individuals with disabilities or special needs. Three additional areas are also important:

- Student knowledge of learning targets and the nature of the assessments prior to instruction (e.g., knowing what will be tested, how it will be graded, scoring criteria, anchors, exemplars, and examples of performance).
- Student prerequisite knowledge and skills, including test-taking skills.
- Avoiding stereotypes.

**Good assessments use multiple methods.**

Assessment that is fair, leading to valid inferences with a minimum of error, is a series of measures that show student understanding through multiple methods. A complete picture of what students understand and can do is put together in pieces comprised by different approaches to assessment. While testing experts and testing companies stress that important decisions should not be made on the basis of a single test score, some educators at the local level, and some (many?) politicians at the state or the national level, seem determined to violate this principle. There is a need to understand the entire range of assessment techniques and methods, with the realization that each has limitations.

**Good assessment is efficient and feasible.**

Teachers and school administrators have limited time and resources. Consideration must be given to the efficiency of different approaches to assessment, balancing needs to implement methods required to provide a full understanding with the time needed to develop and implement the methods, and score results. Teacher skills and knowledge are important to consider, as well as the level of support and resources.

**Good assessment appropriately incorporates technology.**

As technology advances and teachers become more proficient in the use of technology, there will be increased opportunities for teachers and administrators to use computer-based techniques (e.g., item banks, electronic grading, computer-adapted testing, computer-based simulations), Internet resources, and more complex, detailed ways of reporting results. There is, however, a danger that technology will contribute to the mindless use of new resources, such as using items on-line developed by some companies without adequate evidence of reliability, validity, and fairness, and crunching numbers with software programs without sufficient thought about weighting, error, and averaging.

To summarize, what is most essential about assessment is understanding how general, fundamental assessment principles and ideas can be used to enhance student learning and teacher effectiveness. This will be achieved as teachers and administrators learn about conceptual and technical assessment concepts, methods, and procedures, for both large-scale and classroom assessments, and apply these fundamentals to instruction.

## Notes:

An earlier version of this paper was presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 24, 2000.

## References

- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Bruner, J. S. (1960). *The process of education*. NY: Vintage Books.
- Calfee, R. C., & Masuda, W. V. (1997). Classroom assessment as inquiry. In G. D. Phye (Ed.) *Handbook of classroom assessment: Learning, adjustment, and achievement*. NY: Academic Press.
- Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.) *Handbook of classroom assessment: Learning, adjustment, and achievement*. NY: Academic Press.
- Code of fair testing practices in education* (1988). Washington, DC: Joint Committee on Testing Practices (American Psychological Association). Available <http://ericae.net/code.htm>
- Code of professional responsibilities in educational measurement* (1995). Washington, DC: National Council on Measurement in Education. Available <http://www.unl.edu/buros/article2.html>
- Ebel, R. L. (1962). Measurement and the teacher. *Educational Leadership*, 20, 20-24.
- Farr, R., & Griffin, M. (1973). Measurement gaps in teacher education. *Journal of Research and Development in Education*, 7(1), 19-28.
- Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway (Ed.), *Testing in the schools*, San Francisco: Jossey-Bass.
- Gullickson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research*, 79(2), 96-100.
- Gullickson, A. R. (1996). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement*, 23(4), 347-354.
- Impara, J. C., & Plake, B. S. (1996). Professional development in student assessment for educational administrators. *Educational Measurement: Issues and Practice*, 15(2), 14-19.
- Mayo, S. T. (1967). Pre-service preparation of teachers in educational measurement. U.S. Department of Health, Education and Welfare. Washington, DC: Office of Education/Bureau of Research.
- McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators*. Thousand Oaks, CA: Corwin Publishing Company. Available [Amazon.com](http://Amazon.com)

McMillan, J. H., & Nash, S. (2000). Teachers' classroom assessment and grading decision making. Paper presented at the Annual Meeting of the National Council of Measurement in Education, New Orleans.

Rogosa, D. (1999). How accurate are the STAR national percentile rank scores for individual students? - An interpretive guide. Palo Alto, CA: Stanford University.

Sanders, J. R., & Vogel, S. R. (1993). The development of standards for teacher competence in educational assessment of students, in S. L. Wise (Ed.), *Teacher training in measurement and assessment skills*, Lincoln, NB: Burros Institute of Mental Measurements.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice*, 10, (1), 3-6.

Shepard, L. A. (2000). The role of assessment in a learning culture. Paper presented at the Annual Meeting of the American Educational Research Association. Available <http://www.aera.net/meeting/am2000/wrap/praddr01.htm>

*Standards for educational and psychological testing* (1999). Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Standards for teacher competence in educational assessment of students. (1990). American Federation of Teachers, National Council on Measurement in Education, National Education Association. Available: <http://www.unl.edu/buros/article3.html>

Stiggins, R. J. (2000). Classroom assessment: A history of neglect, a future of immense potential. Paper presented at the Annual Meeting of the American Educational Research Association.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press, Albany.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass. Available [Amazon.com](http://www.amazon.com)